Contrastive Latent Variable Models for Neural Text Generation (Supplementary material)

Zhiyang Teng[1,2] **Chenhua Chen**[1,2] **Yan Zhang**[3] **Yue Zhang**[1,2]

[1]School of Engineering, Westlake University, China
[2]Institute of Advanced Technology, Westlake Institute for Advanced Study, China
[3]National University of Singapore

1 DATA STATISTICS FOR LANGUAGE MODELING

Table 1: Statistics of datasets for language modeling.

DATASET	#SENT	MIN_l	MAX_l	AVG_l	#WORD
PTB-TRAIN	41,931	2	82	21.2	887,384
PTB-VALID	3,357	2	74	21.0	70,377
PTB-TEST	3,756	2	77	20.9	78,664
Yahoo-TRAIN	100,000	20	200	78.7	7,872,281
Yahoo-VALID	10,000	20	200	79.1	790,680
Yahoo-TEST	10,000	20	200	78.9	788,673
Yelp-TRAIN	100,000	20	201	96.0	9,603,135
Yelp-VALID	10,000	20	200	96.1	961,392
Yelp-TEST	10,000	20	200	95.7	956,556

Table 1 shows the data statistics for three benchmarks for language modeling. PTB-TRAIN contains about 41K sentence, with an average length of 21.2 words. Both Yahoo-Train and Yelp-Train contain 100K sentences with an average length of more than 78 words.

2 DIALOG RESPONSE GENERATION ON DAILYDIALOG

Language modeling are typical unconditional text generation tasks. Both text summarization and data2text are conditional text generation tasks. We use dialogue response generation on DailyDialog [Li et al., 2017] as an example to investigate our model for conditional text generation. For this task, a dialogue context is given and the goal is to generate an utterance as the response according to the dialogue context. The following Table 2 shows the results.

Table 2: Conditional text generation results on DailyDialog.

MODEL	BLEU-R	BLEU-P	BLEU-F1
SEQGAN [YU ET AL., 2017]	27.0	27.0	27.0
CONDITIONALVAE [ZHAO ET AL., 2017]	26.5	22.2	24.2
IVAE_{MI} [FANG ET AL., 2019]	**35.5**	23.9	28.5
OURS	30.6	**28.6**	**29.6**

Table 2 shows that the conditional VAE gives the worst results. SeqGAN is better than the conditionalVAE. iVAEmi gives better results than simple conditional VAE and SeqGAN. Our model gives better BLEU-P and BLEU-F1 scores than iVAEmi.

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

This suggests that our model can be beneficial for dialogue response generation. We believe that our model can also work for other conditional text generation tasks including text summarization and data2text. We will consider investigating such settings in future.

References

Le Fang, Chunyuan Li, Jianfeng Gao, Wen Dong, and Changyou Chen. Implicit deep latent variable models for text generation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3946–3956, November 2019.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. Dailydialog: A manually labelled multi-turn dialogue dataset. *arXiv preprint arXiv:1710.03957*, 2017.

Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 2852–2858. AAAI Press, 2017.

Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 654–664, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1061. URL `https://aclanthology.org/P17-1061`.